

ERROR PROBABILITIES FOR HALFSPACE DEPTH

MICHAEL A. BURR AND ROBERT FABRIZIO

ABSTRACT. Data depth functions are a generalization of one-dimensional order statistics and medians to real spaces of dimension greater than one; in particular, a data depth function quantifies the centrality of a point with respect to a data set or a probability distribution. One of the most commonly studied data depth functions is halfspace depth. It is of interest to computational geometers because it is highly geometric, and it is of interest to statisticians because it shares many desirable theoretical properties with the one-dimensional median. As the sample size increases, the halfspace depth for a sample converges to the halfspace depth for the underlying distribution, almost surely. In this paper, we use the geometry of halfspace depth to improve the explicit bounds on the rate of convergence.

1. INTRODUCTION

Data depth functions generalize order statistics and the median in one dimension to higher dimensions; in particular, they provide a quantitative estimate for the centrality of a point relative to a data set or a probability distribution (see [12] and [11] for surveys). For these functions, larger values at a point indicate that the point is deeper or more central with respect to a data set or distribution. The point with largest depth is defined to be the *median* of a data set or distribution. For data sets, data depth functions are typically defined in terms of the geometry of the data, and they reflect the geometric combinatorics of the data set.

Halfspace depth is a data depth measure that was first introduced by Hodges [10] and Tukey [15]. Halfspace depth has attracted the interest of computational geometers because of its strong geometric properties (see, for example, [1] and [3]) and is of interest to statisticians because it shares many theoretical properties with the one-dimensional median [17]. We recall the definition of halfspace depth for distributions and data sets, briefly using the notation $\mathcal{H}(q)$ for the set of halfspaces in \mathbb{R}^d containing $q \in \mathbb{R}^d$.

Definition 1.1. Let X be an \mathbb{R}^d -valued random variable. For a point $q \in \mathbb{R}^d$, the *halfspace depth* of q with respect to X is the following minimum over all halfspaces H of \mathbb{R}^d containing q :

$$\text{HD}(q; X) = \min_{H \in \mathcal{H}(q)} \Pr(X \in H).$$

Let $X^{(n)} = (X_1, \dots, X_n)$ be a finite sample of n points in \mathbb{R}^d . For a point $q \in \mathbb{R}^d$, the *halfspace depth* of q with respect to $X^{(n)}$ is the following minimum over all halfspaces H of \mathbb{R}^d containing q :

$$\text{HD}(q; X^{(n)}) = n^{-1} \cdot \min_{H \in \mathcal{H}(q)} \#\{X^{(n)} \cap H\}.$$

In [8], the authors prove that for iid X_1, X_2, \dots random variables equal in distribution to X , as $n \rightarrow \infty$

$$\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \rightarrow 0 \quad \text{a.s.}$$

Date: May 17, 2016.

Key words and phrases. Data depth, Halfspace depth, Convergence, Gilvenko-Cantelli.

This work was partially supported by grants from the Simons Foundation (#282399 to Michael Burr) and the NSF (#CCF-1527193).

In this paper, we use the geometry of halfspace depth to improve explicit bounds on the rate of convergence of this limit. In particular, we show:

Theorem 1.2. Let X be an \mathbb{R}^d -valued random variable (obeying certain Lipschitz continuity conditions defined below) and let (X_1, X_2, \dots) be a sequence of iid random variables, equal in distribution to X . Fix $\varepsilon > 0$. Then, for n sufficiently large, there is a constant C such that

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - Cn^{\frac{3}{2}(d-1)}e^{-2n\varepsilon^2}.$$

This theorem represents an improvement by a factor of $n^{\frac{1}{2}d + \frac{7}{2}}$ over the previous bound. The remainder of this paper is organized as follows: In Section 2, we provide the necessary background on halfspace depth and discuss what is known about its convergence. In Section 3, we quote results from computational geometry and probability theory that we need in the main results. In this section, we also provide the notation for the Lipschitz continuity conditions mentioned in Theorem 1.2. In Section 5, we illustrate the main theorem with a few examples and special cases, and, finally, we conclude in Section 6.

2. BACKGROUND ON HALFSPACE DEPTH AND ITS CONVERGENCE

We begin this section by presenting equivalent formulations of halfspace depth and discuss the geometric properties and notations that are used in the remainder of this paper. In the second part of this paper, we recall probabilistic estimates on the error of the empirical measure of a sample and apply these estimates to halfspace depth.

2.1. Equivalent Definitions of Halfspace Depth. Halfspace depth is a commonly studied and used data depth function because it is simple to define, can be computed efficiently, and satisfies all of the desirable properties for a data depth function as defined in [17]. In the context of this paper, we recall two equivalent formulations of halfspace depth.

Observe, first, that a parallel translation of the boundary of a halfspace in the direction of the halfspace only decreases the measure or number of points in the halfspace. Therefore, in the definition of halfspace depth, it is enough to only consider halfspaces where q lies in the boundary of H . Therefore, we can rewrite the definition of halfspace depth as follows:

Proposition 2.1. Let X be an \mathbb{R}^d -valued random variable and $X^{(n)} = (X_1, \dots, X_n)$ a finite sample of n points in \mathbb{R}^d . For a point $q \in \mathbb{R}^d$, the halfspace depth of q with respect to X or $X^{(n)}$ is the following minimum over all halfspaces H of \mathbb{R}^d whose boundary ∂H contains q :

$$\text{HD}(q; X) = \min_{\partial H \ni q} \Pr(X \in H) \quad \text{and} \quad \text{HD}(q; X^{(n)}) = n^{-1} \cdot \min_{\partial H \ni q} \#\{X^{(n)} \cap H\}.$$

Next, we reinterpret this definition in terms of projections onto one-dimensional subspaces of \mathbb{R}^d as in [8]; this allows us to interpret a high-dimensional problem as a collection of one-dimensional problems. For $q \in \mathbb{R}^d$, the set of halfspaces H with $q \in \partial H$ can be parametrized by points in the $(n-1)$ -dimensional sphere, S^{n-1} . More precisely, each direction in S^{n-1} corresponds to a vector v which describes a halfspace H as follows: the bounding hyperplane ∂H of the halfspace passes through q and has normal v and the halfspace opens in the direction away from v .

We now take this idea and look at it from a different perspective; instead of fixing the point q and considering all halfspaces whose boundary passes through q , we focus on the parameterization given by S^{d-1} . In order to simplify the discussion, we use the following notation:

Notation 2.2. Let $\theta \in S^{d-1}$, define u_θ to be the vector in \mathbb{R}^d pointing in the direction of θ , ℓ_θ to be the line through the origin in the direction of θ , and $\pi_\theta : \mathbb{R}^d \rightarrow \ell_\theta$ to be the orthogonal projection onto ℓ_θ . Moreover, for a point $p \in \mathbb{R}^d$, define $d_\theta(p) = \langle p, u_\theta \rangle$ (where $\langle \cdot, \cdot \rangle$ is the standard inner product in \mathbb{R}^d) to be the signed length of p in the direction of u_θ . In other words, $\pi_\theta(p) = d_\theta(p)u_\theta$.

Using this notation, we define probability distributions and finite samples on \mathbb{R} corresponding to each direction in S^{d-1} .

Notation 2.3. Let X be an \mathbb{R}^d -valued random variable and $X^{(n)} = (X_1, \dots, X_n)$ a finite sample of n points in \mathbb{R}^d . For each $\theta \in S^{n-1}$, X_θ is the \mathbb{R} -valued random variable $d_\theta(X)$, and F_θ is the cdf for this variable, i.e., $F_\theta(t) = \Pr(X_\theta \leq t)$. Similarly, $F_{n,\theta}$ is the empirical cdf for the points of $X^{(n)}$ in the direction of θ , i.e., $F_{n,\theta}(t) = n^{-1} \cdot \#\{i : d_\theta(X_i) \leq t\}$.

For each t and θ , there is a halfspace $H_{\theta,t}$ such that $F_\theta(t) = \Pr(X \in H_{\theta,t})$ and $F_{n,\theta}(t) = n^{-1} \cdot \#\{X^{(n)} \cap H_{\theta,t}\}$. In particular, $H_{\theta,t}$ is the halfspace whose bounding hyperplane passes through the point $t \cdot u_\theta$, whose bounding hyperplane has normal u_θ , and the halfspace opens in the direction of $-u_\theta$. Since the point q lies in the bounding hyperplane for $H_{\theta,t}$ iff $d_\theta(q) = t$, we can reinterpret Proposition 2.1 as follows (see Figure 1):

Proposition 2.4. Let X be an \mathbb{R}^d -valued random variable and $X^{(n)} = (X_1, \dots, X_n)$ a finite sample of n points in \mathbb{R}^d . For a point $q \in \mathbb{R}^d$, the halfspace depth of q with respect to X or $X^{(n)}$ is the following minimum over directions $\theta \in S^{d-1}$:

$$\text{HD}(q; X) = \min_{\theta \in S^{d-1}} F_\theta(d_\theta(q)) \quad \text{and} \quad \text{HD}(q; X^{(n)}) = \min_{\theta \in S^{d-1}} F_{n,\theta}(d_\theta(q)).$$

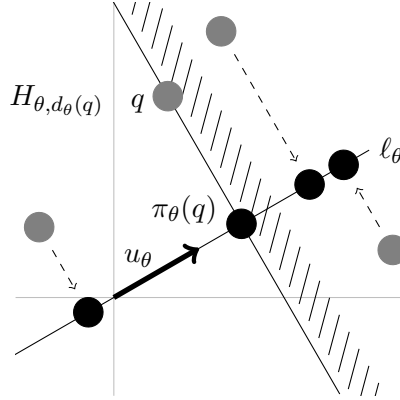


FIGURE 1. The number of points in the halfspace $H_{\theta, d_\theta(q)}$ (the unshaded halfspace) equals the number of points such that $d_\theta(X_i) \leq d_\theta(q)$. In the diagram, these are the points such that $\pi_\theta(X_i)$ is to the left of $\pi_\theta(q)$. A similar statement and diagram can be made for a \mathbb{R}^d -valued random variable X .

These equivalent formulations for halfspace depth illustrate that the halfspace depth of a point q measures how extreme q is under all orthogonal projections onto one-dimensional subspaces. This formulation gives rise to the common description of the halfspace depth of a point q relative to a sample $X^{(n)}$ as the smallest fraction of points that must be removed from the sample so that q is outside the convex hull of the remaining sample points. The formulation of halfspace depth in Proposition 2.4 gives the key approach that is exploited in the main result of this paper.

2.2. Prior Convergence Estimates. Suppose that X is \mathbb{R}^d -valued random variable and that (X_1, X_2, \dots) is a sequence of iid random variables equal in distribution to X . In [8], the authors prove that for all $q \in \mathbb{R}^d$, $\text{HD}(q; X^{(n)}) \rightarrow \text{HD}(q; X)$ almost surely as $n \rightarrow \infty$. This result is proved by observing that the collection of all halfspaces in \mathbb{R}^d satisfy the Glivenko-Cantelli property (for additional details, see [14]), so that, uniformly for all halfspaces H , $n^{-1} \cdot \#\{X^{(n)} \cap H\} \rightarrow \Pr(X \in H)$ a.s. as $n \rightarrow \infty$.

The convergence given by the Glivenko-Cantelli property can be strengthened by observing that the set of halfspaces are a Vapnik-Červonenkis class. In particular, it is shown in [9] that halfspaces in \mathbb{R}^d cannot shatter sets of size $d+2$ (and convex sets of size $d+1$ can be shattered). We write $m(n)$ for the maximum number of subsets formed by intersecting finite samples of size n with halfspaces in \mathbb{R}^d . Then, in [16] and [5], the following result is proved:

Proposition 2.5 (See [14, Chapter 26]). Let \mathcal{H} be the set of all halfspaces in \mathbb{R}^d and suppose that $\varepsilon > 0$. Define m to be the function as described above. Then, for n sufficiently large,

- (1) $\Pr(\sup_{H \in \mathcal{H}} |n^{-1} \cdot \#\{X^{(n)} \cap H\} - \Pr(X \in H)| \geq \varepsilon) \leq 4m(2n)e^{-n\varepsilon^2/8}.$
- (2) $\Pr(\sup_{H \in \mathcal{H}} |n^{-1} \cdot \#\{X^{(n)} \cap H\} - \Pr(X \in H)| \geq \varepsilon) \leq 4m(n^2)e^{-2n\varepsilon^2}.$

Moreover, $m(r) \leq \frac{3}{2} \cdot \frac{r^{d+1}}{(d+1)!}.$

We improve these bounds on the error by decreasing the exponent or the degree of the polynomial coefficient.

3. ADDITIONAL TOOLS

Since data depth combines discrete geometry with probability and statistics, our main result requires results from both of these fields. In this section, for convenience, we collect a few additional theorems and notations that are used in the remainder of the paper.

3.1. Spherical Covering. As observed in Section 2.1, the halfspace depth is related to the directions in a $(d-1)$ -dimensional sphere. In our main result, we consider small neighborhoods on a $(d-1)$ -dimensional sphere. The following result indicates how many of these neighborhoods are necessary.

Lemma 3.1 ([2, Corollary 1.2]). For any $0 < \psi < \arccos(d^{-1/2})$, there is an absolute constant C_2 such that the $(d-1)$ -dimensional unit sphere can be covered by

$$C_2 \frac{\cos \psi}{\sin^{d-1} \psi} (d-1)^{\frac{3}{2}} \ln(1 + (d-1) \cos^2 \psi) \leq C_2 \left(\frac{\sqrt{d}}{\psi} \right)^{d-1} (d-1)^{\frac{3}{2}} \ln(d)$$

spherical balls of radius ψ (i.e., the radius on the surface of the sphere is ψ).

The final inequality in this lemma follows from the facts that $\cos(\psi) \leq 1$ and for $0 < \psi < \arccos(d^{-1/2})$, $\sin(\psi) \geq \psi \cdot \cos(\psi) > d^{-1/2}\psi$. The \sqrt{d} could be replaced by a constant by using a better bound on $\sin(x)$, such as $\sin(x) \geq x - \frac{x^3}{3!}$. We leave the details to the interested reader.

3.2. The Dvoretzky-Kiefer-Wolfowitz Inequality. In the one-dimensional case, the bounds in Proposition 2.5 can be improved with the Dvoretzky-Kiefer-Wolfowitz inequality.

Lemma 3.2 (See, for example, [6]). Let X be an \mathbb{R} -valued random variable, and let (X_1, X_2, \dots) be a sequence of iid random variables equal in distribution to X . Let F be the cdf of X , and let F_n be the empirical cdf, i.e., $F_n(t) = n^{-1} \cdot \#\{i \leq n : X_i \leq t\}$. For each $\varepsilon > 0$,

$$\Pr(\sup_{t \in \mathbb{R}} |F(t) - F_n(t)| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

In this paper, we use the Dvoretzky-Kiefer-Wolfowitz inequality on one-dimensional projections, as in Proposition 2.4, in order to improve the bound on the convergence rate of the sample halfspace depth.

3.3. Lipschitz and Tail Behavior. In this section, we define a few conditions on the probability distributions that we use for our main result. Let X be an \mathbb{R}^d -valued random variable with $d > 1$.

Definition 3.3. We say that an \mathbb{R}^d -valued random variable X *decays quickly* if there is some $\lambda > 0$ such that $\Pr(|X| > R) = O\left(R^{3d-5} \cdot e^{-\lambda \frac{R^2}{2}}\right)$. In other words, there exists a constant $C_1 > 0$ such that for $R > 1$, $\Pr(|X| > R) < C_1 \cdot R^{3d-5} \cdot e^{-\lambda \frac{R^2}{2}}$. We say that λ is the *decay rate* of X .

For example, the multivariate normal with the origin the mean and covariance matrix I decays exponentially with $\lambda = 1$.

Definition 3.4. We say that X is *Lipschitz continuous in projection* if there is a constant $L_\pi > 0$ such that for all $\theta \in S^{d-1}$, F_θ is Lipschitz continuous with Lipschitz constant L_π . We say that X is *radially Lipschitz continuous* if there is a constant L_θ such that for any fixed t , the function $F_\theta(t)$ is Lipschitz continuous as a function of θ on the unit circle with Lipschitz constant L_θ . We say that L_π is the *projection Lipschitz constant* and L_θ is the *radial Lipschitz constant* for X .

For example, radially symmetric distributions about the origin have $L_\theta = 0$. These two Lipschitz constants indicate that as θ and t change, $F_\theta(t)$ changes continuously. In other words, these two constants can be used to bound the difference in the probability between two halfspaces and to show that this difference varies continuously.

4. MAIN RESULT

We begin this section by highlighting the difference between the previous bounds and our approach. In Proposition 2.5, the convergence rates are computed over all halfspaces in \mathbb{R}^d , and the goal is to find a bound on

$$\Pr\left(\sup_{H \in \mathcal{H}} |n^{-1} \cdot \#\{X^{(n)} \cap H\} - \Pr(X \in H)| \geq \varepsilon\right).$$

In our approach, however, the goal is to find bounds on the family of one-dimensional cdfs F_θ and $F_{n,\theta}$. In particular, in our main result, we find a uniform upper bound on the following probability:

$$\Pr\left(\sup_{\substack{t \in \mathbb{R} \\ \theta \in S^{d-1}}} |F_\theta(t) - F_{n,\theta}(t)| \geq \varepsilon\right).$$

Then, by applying Proposition 2.4, when $|F_\theta(t) - F_{n,\theta}(t)| < \varepsilon$ for all t and θ , it follows that for all $q \in \mathbb{R}^d$, $|\text{HD}(q; X) - \text{HD}(q; X^{(n)})| < \varepsilon$. The advantage of this approach is that by considering only one-dimensional objects, we can use the improved bounds of the Dvoretzky-Kiefer-Wolfowitz inequality.

Throughout the remainder of this section, we assume that the \mathbb{R}^d -valued random variable X with $d > 1$ has decay rate λ with constant C_1 , projection Lipschitz constant L_π , and radial Lipschitz constant L_θ . Also, let (X_1, X_2, \dots) be a sequence of iid random variables equal in distribution to X .

Our proof follows the following steps: (1) First, we quantify how small changes in θ affect the cdfs F_θ and $F_{n,\theta}$. (2) Second, we use this description on F_θ and $F_{n,\theta}$ to argue that it is sufficient to examine finitely many θ 's instead of the infinitely many possible θ 's in S^{d-1} . (3) Finally, we carefully choose a few parameters in order to achieve a bound which improves upon the bounds achieved in Proposition 2.5.

For our first step, we begin by collecting, for later convenience, a consequence of the Lipschitz condition, and then provide a geometric description of the behavior of d_θ as θ changes. For notational convenience, for $\theta, \varphi \in S^{d-1}$, we write $|\theta - \varphi|$ for the spherical distance between them, i.e., the central angle between the two points.

Observation 4.1. For all $\theta, \varphi \in S^{d-1}$ and $t \in \mathbb{R}$, $|F_\theta(t) - F_\varphi(t)| \leq L_\theta |\theta - \varphi|$.

Lemma 4.2. Let $\theta, \varphi \in S^{d-1}$ and $x \in \mathbb{R}^d$. Then, $|d_\theta(x) - d_\varphi(x)| \leq \|x\| \cdot |\theta - \varphi|$.

Proof. Observe first that $|d_\theta(x) - d_\varphi(x)| = |\langle x, u_\theta - u_\varphi \rangle| \leq \|x\| \cdot \|u_\theta - u_\varphi\|$. Note that the vector $u_\theta - u_\varphi$ is the chord of an arc on the great circle between θ and φ ; since chords are shorter than the arcs they cut off, and the length of the arc is $|\theta - \varphi|$, the result follows. \square

Moreover, Lemma 4.2 allows us to bound $F_{n,\varphi}$ in terms of $F_{n,\theta}$. More precisely:

Corollary 4.3. Fix $R > 0$ and $\theta, \varphi \in S^{d-1}$. Suppose that for all $i \leq n$, $\|X_i\| \leq R$. Then,

$$F_{n,\theta}(t - R|\theta - \varphi|) \leq F_{n,\varphi}(t) \leq F_{n,\theta}(t + R|\theta - \varphi|).$$

Proof. This result follows from Lemma 4.2, because we know that if $d_\varphi(X_i) \leq t$, then $d_\theta(X_i) \leq t + R \cdot |\theta - \varphi|$, and if $d_\theta(X_i) \leq t - R \cdot |\theta - \varphi|$, then $d_\varphi(X_i) \leq t$. \square

With Observation 4.1 and Corollary 4.3 in hand, we know how the distributions F_θ and $F_{n,\theta}$ vary as θ varies. This leads to the following proposition, which bounds the error between F_θ and $F_{n,\theta}$ as θ varies.

Proposition 4.4. Fix $R > 0$, $\varepsilon > 0$, $\delta > 0$, and $\theta, \varphi \in S^{d-1}$. Suppose that for all $i \leq n$, $\|X_i\| \leq R$. Suppose that $\sup_{t \in \mathbb{R}} |F_\theta(t) - F_{n,\theta}(t)| < \varepsilon \cdot (1 + \delta)^{-1}$. Then,

$$\sup_{t \in \mathbb{R}} |F_\varphi(t) - F_{n,\varphi}(t)| < \varepsilon(1 + \delta)^{-1} + L_\theta|\theta - \varphi| + L_\pi R|\theta - \varphi|.$$

Proof. By Corollary 4.3, we know that

$$F_{n,\theta}(t - R|\theta - \varphi|) - F_\varphi(t) \leq F_{n,\varphi}(t) - F_\varphi(t) \leq F_{n,\theta}(t + R|\theta - \varphi|) - F_\varphi(t).$$

Then, it follows that

$$|F_{n,\varphi}(t) - F_\varphi(t)| \leq \max\{|F_{n,\theta}(t - R|\theta - \varphi|) - F_\varphi(t)|, |F_{n,\theta}(t + R|\theta - \varphi|) - F_\varphi(t)|\}.$$

We proceed to bound the first of the expression in the maximum (the other expression is similar). By the triangle inequality,

$$\begin{aligned} |F_{n,\theta}(t - R|\theta - \varphi|) - F_\varphi(t)| \\ \leq |F_{n,\theta}(t - R|\theta - \varphi|) - F_\theta(t - R|\theta - \varphi|)| + |F_\theta(t - R|\theta - \varphi|) - F_\theta(t)| + |F_\theta(t) - F_\varphi(t)|. \end{aligned}$$

By assumption, the first expression is bounded above by $\varepsilon \cdot (1 + \delta)^{-1}$. By the projection Lipschitz constant, the second expression is bounded above by $L_\pi \cdot R \cdot |\theta - \varphi|$. Finally, by Observation 4.1, the third expression is bounded by $L_\theta \cdot |\theta - \varphi|$. Combining these bounds, the result follows. \square

By choosing $|\theta - \varphi|$ to be sufficiently small, as in the following corollary, we can insist that the errors are all bounded above by ε .

Corollary 4.5. Fix $R > 0$, $\varepsilon > 0$, $\delta > 0$, and $\theta \in S^{d-1}$. Suppose that for all $i \leq n$, $\|X_i\| \leq R$. Let $\varphi \in S^{d-1}$ be such that $|\theta - \varphi| < \varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1}$. Suppose that $\sup_{t \in \mathbb{R}} |F_\theta(t) - F_{n,\theta}(t)| < \varepsilon \cdot (1 + \delta)^{-1}$. Then, $\sup_{t \in \mathbb{R}} |F_\varphi(t) - F_{n,\varphi}(t)| < \varepsilon$.

Corollary 4.5 completes the second main step of the proof and gives us a way to study neighborhoods on S^{d-1} instead of individual projections. In particular, if $\{\theta_1, \dots, \theta_k\}$ satisfy the conditions of Corollary 4.5, then we have estimates on the error in the k spherical balls of radius $\varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1}$ centered at the θ_i 's. We now determine the θ_i 's and use the Dvoretzky-Kiefer-Wolfowitz bound in these directions to get our initial estimate on the error between the halfspace depth of a sample and the underlying distribution. In particular,

Proposition 4.6. Fix $R > 1$, $\varepsilon > 0$, and $\delta > 0$. Suppose that $\varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1} < \arccos(d^{-1/2})$. Then,

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - 2C_2 \left(\frac{(1 + \delta)(L_\theta + L_\pi R)\sqrt{d}}{\varepsilon \delta} \right)^{d-1} (d-1)^{\frac{3}{2}} \ln(d) e^{-2n\varepsilon^2(1+\delta)^{-2}} - C_1 n R^{3d-5} e^{-\lambda \frac{R^2}{2}}. \quad (1)$$

Proof. Observe first, that, in each of the statements above, we require that for all $i \leq n$, $\|X_i\| \leq R$. By the exponential decay property, for each i , the probability of this occurring is bounded below by $1 - C_1 R^{3d-5} e^{-\lambda \frac{R^2}{2}}$.

Next, observe that if $\theta \in S^{d-1}$ is such that $\sup_{t \in \mathbb{R}} |F_\theta(t) - F_{n,\theta}(t)| < \varepsilon \cdot (1 + \delta)^{-1}$, then for all $\varphi \in S^{d-1}$ within a spherical ball of radius $\varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1}$ satisfy the conditions of Corollary 4.5. Therefore, we cover the $(d-1)$ -dimensional sphere with balls of radius $\varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1}$. By Lemma 3.1, $C_2 \cdot \left((1 + \delta)(L_\theta + L_\pi R)\sqrt{d} \right)^{d-1} \cdot (\varepsilon \delta)^{1-d} \cdot (d-1)^{\frac{3}{2}} \cdot \ln(d)$ such spherical balls are required. Fix θ to be the center of one of these spherical balls. Applying the Dvoretzky-Kiefer-Wolfowitz inequality, Lemma 3.2, it follows that

$$\Pr(\sup_{t \in \mathbb{R}} |F_\theta(t) - F_{n,\theta}(t)| \leq \varepsilon(1 + \delta)^{-1}) \geq 1 - 2e^{-2n\varepsilon^2(1+\delta)^{-2}}.$$

Since the spherical balls cover S^{d-1} , for any $\varphi \in S^{d-1}$, there is some spherical ball with center θ such that $|\theta - \varphi| < \varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot (L_\theta + L_\pi \cdot R)^{-1}$. Hence, by Corollary 4.5, it follows that $\sup_{t \in \mathbb{R}} |F_\varphi(t) - F_{n,\varphi}(t)| < \varepsilon$.

Finally, using the fact that for $a, b > 0$, $(1 - a)(1 - b) > 1 - (a + b)$, and that there are n sample points and $C_2 \cdot \left((1 + \delta)(L_\theta + L_\pi R)\sqrt{d} \right)^{d-1} \cdot (\varepsilon \delta)^{1-d} \cdot (d-1)^{\frac{3}{2}} \cdot \ln(d)$ spherical balls, the result follows. \square

We now simplify the expression above by eliminating the parameters R and δ . Since the bound in Inequality (1) has two exponentials, we can choose $R = \varepsilon \cdot \frac{2\sqrt{n}}{\sqrt{\lambda}(1+\delta)}$ to equate the exponentials. This choice results in the following corollary:

Corollary 4.7. Fix $\varepsilon > 0$ and $\delta > 0$. Suppose that $\varepsilon \cdot \delta \cdot (1 + \delta)^{-1} \cdot \left(L_\theta + L_\pi \cdot \varepsilon \cdot \frac{2\sqrt{n}}{\sqrt{\lambda}(1+\delta)} \right)^{-1} < \arccos(d^{-1/2})$ and $2\varepsilon\sqrt{n} > \sqrt{\lambda}(1 + \delta)$. Then,

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - \left(2C_2 \left(\frac{(L_\theta \sqrt{\lambda}(1 + \delta) + 2L_\pi \sqrt{n}\varepsilon)\sqrt{d}}{\varepsilon \delta \sqrt{\lambda}} \right)^{d-1} (d-1)^{\frac{3}{2}} \ln(d) + C_1 \left(\frac{\varepsilon}{\sqrt{\lambda}(1 + \delta)} \right)^{3d-5} n^{\frac{3}{2}(d-1)} \right) e^{-2n\varepsilon^2(1+\delta)^{-2}}. \quad (2)$$

Finally, we choose $\delta = n^{-1}$ in order to eliminate δ . Additionally, this choice causes the exponential to simplify to $e^{-2n\varepsilon^2}$. More precisely,

Theorem 4.8. Fix $\varepsilon > 0$ and suppose that $\varepsilon \cdot (n+1)^{-1} \cdot \left(L_\theta + L_\pi \cdot \varepsilon \cdot \frac{2n^{3/2}}{\sqrt{\lambda(n+1)}} \right)^{-1} < \arccos(d^{-1/2})$ and $2\varepsilon n^{\frac{3}{2}} > \sqrt{\lambda}(n+1)$. Then,

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - \left(2C_2 \left(\frac{(L_\theta \sqrt{\lambda}(n+1) + 2L_\pi n^{\frac{3}{2}} \varepsilon) \sqrt{d}}{\varepsilon \sqrt{\lambda}} \right)^{d-1} (d-1)^{\frac{3}{2}} \ln(d) + C_1 \left(\frac{\varepsilon n}{\sqrt{\lambda}(n+1)} \right)^{3d-5} n^{\frac{3}{2}(d-1)} \right) e^4 e^{-2n\varepsilon^2}. \quad (3)$$

Proof. By setting $\delta = n^{-1}$ in Inequality (2) and expanding the exponential, the bound in Corollary 4.7 becomes

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - \left(2C_2 \left(\frac{(L_\theta \sqrt{\lambda}(n+1) + 2L_\pi n^{\frac{3}{2}} \varepsilon) \sqrt{d}}{\varepsilon \sqrt{\lambda}} \right)^{d-1} (d-1)^{\frac{3}{2}} \ln(d) + C_1 \left(\frac{\varepsilon n}{\sqrt{\lambda}(n+1)} \right)^{3d-5} n^{\frac{3}{2}(d-1)} \right) e^{-2\varepsilon^2 \left(n - \frac{2n^2+n}{n^2+2n+1} \right)}.$$

Since ε is a difference between two quantities whose values are between 0 and 1, only $\varepsilon \leq 1$ has content. Moreover, the quotient in the exponential is decreases to -2 for n positive; therefore, the exponent is bounded above by $-2n\varepsilon^2 + 4$, and the result follows. \square

We summarize Theorem 4.8 for n sufficiently large.

Corollary 4.9. Fix $\varepsilon > 0$. Then, for n sufficiently large, there is a constant C depending on λ , ε , C_1 , C_2 , d , and L_π such that

$$\Pr \left(\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| \leq \varepsilon \right) \geq 1 - C n^{\frac{3}{2}(d-1)} e^{-2n\varepsilon^2}.$$

We can compare this result with the known convergence rates in Proposition 2.5. The exponential in Inequality (1) is much larger (the exponent is a negative number with smaller magnitude) than what appears in Corollary 4.9, so our bound is tighter. On the other hand, the best coefficient in Inequality (2) is n^{2d+2} , which is a factor of $n^{\frac{1}{2}d+\frac{7}{2}}$ times larger than our bound.

5. EXAMPLES

In this section, we apply Theorem 4.8 to several examples. We take special care to develop explicit bounds on the Lipschitz constants whenever possible.

Example 5.1. Consider the case where X is distributed according to an elliptically symmetric distribution with pdf $f(x) = \det(\Sigma)^{-\frac{1}{2}} \psi((x - \mu)^T \Sigma^{-1}(x - \mu))$ where x and μ are in \mathbb{R}^d , Σ is a positive definite symmetric matrix, and $\psi : [0, \infty) \rightarrow [0, \infty)$ (see, for example, [4]). Since Σ is positive definite, we can decompose $\Sigma = QD^2Q^T$ where the entries of Q are the orthogonal eigenvectors of Σ and D is a diagonal matrix whose entries are the square roots of the eigenvalues of Σ .

Consider the affine transformation $Y = D^{-1}Q^T(X - \mu)$; under this transformation, the pdf of Y becomes $\psi(\sum y_i^2)$. Recall that since halfspace depth is invariant under affine transformations (see, for example, [17]). Therefore,

$$\sup_{q \in \mathbb{R}^d} |\text{HD}(q; X) - \text{HD}(q; X^{(n)})| = \sup_{q \in \mathbb{R}^d} |\text{HD}(q; Y) - \text{HD}(q; Y^{(n)})|,$$

and we can study the convergence of X by studying the convergence of Y . Throughout the remainder of this example all statements refer to Y .

We suppose, additionally, that ψ has the following properties:

- (1) $\int_0^\infty r^{d-1}\psi(r^2)dr = (\text{Vol}_{d-1}(S^{d-1}))^{-1}$. This condition guarantees that the integral of the pdf for Y over \mathbb{R}^d is 1.
- (2) There exists a $\lambda > 0$ such that $\int_R^\infty r^{d-1}\psi(r^2)dr = O(R^{3d-5}e^{-\lambda\frac{R^2}{2}})$. This condition guarantees that X has a decay rate of λ .
- (3) The function $\Psi(t) = \int_{\mathbb{R}^{d-1}} \psi(t^2 + x_1^2 + \dots + x_{d-1}^2)dx_1 \dots dx_{d-1}$ is bounded. This function is the derivative of $F_{e_d}(t)$, where e_d is the north pole of S^{d-1} . The upper bound on this function is a Lipschitz constant for $F_{e_d}(t)$. When Conditions (1) and (2) hold, this is not a strong condition; for example, it holds when ψ is bounded.

Like all distributions which are spherically symmetric about the origin, for Y , $L_\theta = 0$. Moreover, by Condition (2), Y has a decay rate of λ . Finally, since this is a spherically symmetric distribution, $F_\theta(t)$ is independent of θ . By Condition 3, $F'_{e_d}(t) = \int_{\mathbb{R}^{d-1}} \psi(t^2 + x_1^2 + \dots + x_{d-1}^2)dx_1 \dots dx_{d-1}$ is bounded. Therefore, the projection Lipschitz constant is $L_\pi = \sup_{t \in \mathbb{R}} \left| \int_{\mathbb{R}^{d-1}} \psi(t^2 + x_1^2 + \dots + x_{d-1}^2)dx_1 \dots dx_{d-1} \right|$.

In particular cases, we can derive more precise bounds and constants.

Example 5.2. Consider the case where X is distributed according to a non-degenerate normal distribution in \mathbb{R}^d with mean μ and (positive definite) covariance matrix Σ . By applying Example 5.1, it is enough to consider the standard normal distribution in \mathbb{R}^d centered at the origin with covariance matrix I . In this case, the pdf of X is $(2\pi)^{-d/2}e^{-\frac{1}{2}\sum x_i^2}$. Moreover, via spherical integration it follows that $\Pr(|X| > R) = O(R^{d-2}e^{-\frac{R^2}{2}})$, so the decay rate of X is 1. Additionally, for any θ , F_θ is the cdf of a standard normal distribution in one variable, since the Lipschitz constant for F_θ is bounded by the absolute value of the derivative, $L_\pi = \frac{1}{\sqrt{2\pi}}$.

The computation for C_1 in Definition 3.3 is more technical (not particularly interesting); we observe that since $\Pr(|X| > R) = O(R^{d-2}e^{-\frac{R^2}{2}})$, we could, in fact, replace the $3d - 5$ by $d - 2$ in Definition 3.3 and subsequent computations to achieve a better bound with C_1 appearing only in a lower order term. Therefore, we choose to ignore the C_1 term. We leave the details to the interested reader.

Example 5.3. For the two-dimensional normal, we can say even more. More precisely, assume that X is distributed according to a bivariate normal centered at the origin with covariance I . The bounds above apply, but can be made even sharper. For example, we can cover a circle using at most $\frac{\pi}{w} + 1$ intervals of width $2w$. Moreover, $\Pr(|X| > R) = e^{-\frac{1}{2}R^2}$. Therefore, in this case, the bound in Theorem 4.8 has a smaller constant, and, explicitly, the bound becomes

$$1 - \left(2\sqrt{2\pi}n^{\frac{3}{2}} + n + 2\right)e^4e^{-2n\epsilon^2}.$$

Additionally, in the two-dimensional case, the function m for Proposition 2.5 can be computed explicitly. The largest number of subsets of a set of size n occurs when the points of n are in convex position. In this case, there are $n^2 - n + 2$ subsets formed from intersections with halfspaces, so $m(r) = r^2 - r + 2$. Even with this smaller degree polynomial for m , our bounds are still an improvement by a factor of \sqrt{n} .

6. CONCLUSION

The results in this paper illustrate how, using the geometry and topology of halfspace depth and \mathbb{R}^d , one can achieve better convergence bounds for the sample version of halfspace depth, as compared to general Glivenko-Cantelli bounds. With improved bounds in this paper, we have improved estimates on the quality of the halfspace median statistic (see [12] and [1]), which is

applicable in statistics. The approach in this paper is related to the ideas of the projection pursuit in [7] and [13]; it is possible that incorporating such techniques may further improve the convergence rates of halfspace depth; we leave such improvements as future work.

The authors would like to thank their colleagues at Clemson University, in particular, Billy Bridges, Brian Fralix, Peter Kiessler, and June Luo for their constructive feedback on earlier versions of this work.

REFERENCES

- [1] Greg Aloupis. Geometric measures of data depth. In *Data depth: robust multivariate analysis, computational geometry and applications*, volume 72 of *DIMACS Series on Discrete Mathematics and Theoretical Computer Science*, pages 147–158. American Mathematical Society, Providence, RI, 2006.
- [2] Károly Böröczky, Jr. and Gergely Wintsche. Covering the sphere by equal spherical balls. In *Discrete and computational geometry*, volume 25 of *Algorithms Combinatorics*, pages 235–251. Springer, Berlin, 2003.
- [3] Michael A. Burr, Eynat Rafalin, and Diane L. Souvaine. Dynamic maintenance of half-space depth for points and contours. Technical Report arXiv:1109.1517 [cs.CG], arXiv, 2011.
- [4] M. A. Chmielewski. Elliptically symmetric distributions: a review and bibliography. *International Statistical Review. Revue Internationale de Statistique*, 49(1):67–74, 1981.
- [5] Luc Devroye. Upper and lower class sequences for minimal uniform spacings. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61(2):237–254, 1982.
- [6] Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York, 2001.
- [7] Persi Diaconis and David Freedman. Asymptotics of graphical projection pursuit. *The Annals of Statistics*, 12(3):793–815, 1984.
- [8] David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992.
- [9] Richard M. Dudley. Balls in \mathbb{R}^k do not cut all subsets of $k + 2$ points. *Advances in Mathematics*, 31(3):306–308, 1979.
- [10] Joseph L. Hodges, Jr. A bivariate sign test. *Annals of Mathematical Statistics*, 26:523–527, 1955.
- [11] Regina Y. Liu. Data depth: center-outward ordering of multivariate data and nonparametric multivariate statistics. In Michael G. Akritas and Dimitris N. Politis, editors, *Recent advances and trends in nonparametric statistics*, pages 155–167. Elsevier B. V., Amsterdam, 2003.
- [12] Regina Y. Liu, Jesse M. Parelus, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.
- [13] Elizabeth S. Meckes. Quantitative asymptotics of graphical projection pursuit. *Electronic Communications in Probability*, 14:176–185, 2009.
- [14] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.
- [15] John W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2, pages 523–531. Canadian Mathematical Congress, Montreal, Que., 1975.
- [16] V. N. Vapnik and A. Ja. Červonenkis. The uniform convergence of frequencies of the appearance of events to their probabilities. *Akademija Nauk SSSR. Teorija Veroyatnostej i ee Primenenija*, 16:264–279, 1971.
- [17] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000.

DEPARTMENT OF MATHEMATICAL SCIENCES, CLEMSON UNIVERSITY, CLEMSON, SC 29634
E-mail address: burr2@clemson.edu

DEPARTMENT OF MATHEMATICAL SCIENCES, CLEMSON UNIVERSITY, CLEMSON, SC 29634
E-mail address: rfabriz@clemson.edu